2014

Evaluating a cognitive tool built to aid decision making using decision making approach as a theoretical framework and using unobtrusive, behavior-based measures for analysis

Ryan Allen Kirk
Iowa State University

Evaluating a cognitive tool built to aid decision making using decision making approach as a theoretical framework and using unobtrusive, behavior-based measures for analysis

by

Ryan Allen Kirk

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Human-Computer Interaction

Program of Study Committee:
Jared Danielson, Major Professor
Stephen Gilbert
Amanda Fales-Williams
Heike Hoffman
Brian Mennecke

Iowa State University

Ames, Iowa

2014

DEDICATION

This work is dedicated to my wife and my family for their continued patience and understanding as I follow my passions for building technologies to help people make decisions. iv

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to thank my committee chair Jared Danielson for his patience and mentorship. I also wish to thank my committee members Stephen Gilbert, Brian Mennecke, Heike Hoffman, and Amanda Fales-Williams for their guidance and stewardship during the course of this research. I also appreciate those who helped facilitate the research project through acting as experts, reviewers, or participants. Finally, I would like to thank Iowa State University, the Human-Computer Interaction program, and the dedicated professors for creating an environment conducive to conducting research. For me this research would not have been possible without the help from these people and without access to these resources.

ABSTRACT

Previously I designed and built an interface with the purpose of augmenting decision making within a particular curricular decision making context. The present study explores the usability of the newly created cognitive tool through analyzing its impact upon facilitating decision making. The introduction discusses how different types of cognitive tools facilitate decision making from a cognitive perspective. The present study examines the newly created dashboard by first breaking it into its five constituent regions. The methods section discusses the hypothesized function and usage patterns of each region. The primary research question was whether these different regions would cause participants to exhibit different exploratory behaviors. Differences in usage patterns between regions, combined with the knowledge of how different cognitive tools function, allowed this study to classify the function of each region of the cognitive tool. This study also considered several secondary factors including participant experience with technology, experience with curricular decision making, spatial acuity, and performance. The primary contribution of this study is a technique that offers researchers increased capabilities to conduct unobtrusive research that quantitatively informs interface design. The next step for this research is to extend these methods to focus upon longer-term research questions.

CHAPTER 1

INTRODUCTION

The ISU College of Veterinary Medicine (ISUCVM) seeks to use data about curricular performance to improve curriculum continuously. Conversations about curricular improvement can be complex for a number of reasons. Data about student learning outcomes and satisfaction are often incomplete and unorganized because data come from many different sources. Additionally, due mainly to the size of the institution, even when data are well organized there are often problems with incongruence between scales and methods used for assessing students.

The problem of making sense of data from many sources is not unique to the veterinary college. Data-driven decision making is difficult in the presence of complex and contradictory information (Driver, J., 2001; Greenwald, 1992; Roediger, H. L., McDermott, K. B., 2000; Bargh, 1992). People's decisions are also shaped by their life experience and, because of this tendency, group decisions can sometimes be incomplete and non-optimal (Bargh, 1992; Medin, D. L., 1989; McClelland, J. L., 2000). In an effort to reduce errors in curricular decision making, the ISUCVM currently uses specially prepared reports. Not only are these reports time consuming to create, but also the capabilities of information visualization using paper media are limited. The ISUCVM needs an information visualization technology that can compensate for the limitations of paper media, for data inconsistencies, and for the tendency towards making cognitive errors.

One of the reasons that the ISUCVM needs an interactive visualization platform is that, like most schools, the ISUCVM is "data rich but information poor" when it comes to using technology to aid in making decisions about curricular change (Wayman, 2005). Education is a field that has traditionally adopted new technologies slowly and is a secondary if not a tertiary market for information technology (Holloway, R.E., 2002). Government policy is also interested in the use of data-driven decision making using formative assessment of student outcomes (Moss and Piety, 2007).

Building a technology is only the first step in creating a system that can support evidence-based decision making. The implementation of new technology can be a complex, socially connected process that changes over time as ideas about how a technology can or should be used change. Communities of individuals use technologies. In these communities, these individual are interrelated both with each other as well as with the technology. Over time, communication about new ideas can converge or diverge in a non-linear fashion. Rogers (2003), writes that the uncertainty amongst users about how to use a new technology reduces over time. Rogers suggests that community adoption leads to the diffusion of new technology. From a practical perspective, this means that there are often unexpected purposes for new technologies. This is especially true for social technologies since the given community defines what effective use of new technology looks like (Townsend and Bennet, 2007).

Given the problems with making accurate decisions, there is a need for an information visualization platform to aid curricular decision makers with making evidence-based evaluations. This dissertation discusses the development and evaluation of a cognitive tool built purposefully to facilitate conversations about curricular change within the College of Veterinary Medicine.

Defining a Cognitive Tool

The human mind excels at making decisions and categorizations amidst disparate information from varying sources across multiple senses. This natural ability is vast but does not come without limitations. Cognitive errors that people make on a daily basis exemplify these limitations. Understand these seemingly random errors as the systematic limitations of human cognitive abilities. Just as healthy individuals have limits on how far or near they can see without aids such as the telescope or microscope, so too do healthy individuals have similar limitations when it comes to making unaided decisions and categorizations. Cognitive tools aid in facilitating decision making and categorization. Salomon defined a cognitive tool as any device that extends or leverages human cognitive capacity (1988). Just as a physical tool, such as a lever, extends human physical capacity, a cognitive tool extends decision making capacity (Salomon, 1988; see also Engle, 2002). Cognitive tools assist people both in problem solving and in learning how to solve problems.

The scope of cognitive tools

Cognitive tools, as with other tools, vary in scope and complexity. Cognitive tools include a wide variety of technologies ranging from the linguistic to the mathematical to the mechanical. As the communities they interact with and the technologies used by those communities continue to change, designers of cognitive tools purposefully re-define their tools. In this way, cognitive tools naturally change to continue to fit their environment. For instance, linguistic ontologies act as cognitive tools when they give a common language and a common conception to problem solving (Mirolli & Parisi, 2009; see also Vygotsky, 2010). The simple act of categorizing data requires an established and agreed-upon nomenclature.

Once data are categorized, mathematical and statistical analyses becomes possible. These mathematical and statistical methods used to improve human understanding and aid in making categorizations that leverage differences within data. The advent of computers brought about an era of automatic implementation of these mathematical and statistical tools. At this point, statistics and probability became openly accessible, allowing every individual to have the capability to make inferences amidst complex data using computers. Similar advances in information visualization techniques leverage human mental abilities to help people find trends in data.

Approaches to using cognitive tools to facilitate human decision making

Despite the changing nature of cognitive tools, what is important is that cognitive tools facilitate purposeful and efficient human thought. The field of Artificial Intelligence (AI) strives to make computers behave more intelligently by making them behave more like humans. In contrast, AI in reverse strives to make humans behave more intelligently by leveraging computer technology to help humans make better decisions (Salomon, 1988). AI in reverse does not mean that people turn into computers through a form of cybernetics, but it does refer to the mind's tendency to adapt and to learn from tool use. Phrased as a design recommendation, this suggests to either present information in a form that the users expect or in a form that is optimal for the display of particular information (Salomon, 1988). When successful, Salomon (1988) suggests that adhering to these design goals will offer three benefits: 1) language acts as a tool, 2) the cognitive tool provides the necessary social scaffolding and creates a zone of proximal development for users, and 3) users will internalize the tool in their own mental operations.

Kim and Reeves (2007) recommended maintaining a distinction between cognitive tools and the users' decision making processes. Interestingly, Kim and Reeves (2007) also recommend viewing the combination of the users and the system as two parts of an integrated system of expertise. Douglas and Schuler (2010) similarly argued that combining disparate parts of a complex socio-technical system into a community network is a form of emergent intelligence. Given this idea, it is possible to improve an entire socio-technical system through facilitating the acquisition and integration of its members' expertise. In this context, a cognitive tool that helps faculty make better curricular decisions would improve the entire curricular development process.

Liu and Bera (2005) used the approach of examining users as part of a system of expertise. They examined how the system supported each of four cognitive functions using a variety of cognitive tools. The cognitive function categories used were (a) cognitive-based, (b) scaffolding-based, (c) attribution-based, and (d) process-based. Liu and Bera categorized their cognitive tools based upon which of the four cognitive functions they supported. They then asked learners to solve problems using those cognitive tools. Learners chose which tool they would use to solve each problem. Based on the tools that learners chose to address different kinds of problems, Liu and Bera proposed that facilitating a certain cognitive function would require a certain cognitive tool.

Cognitive-Based Approach to Facilitating Decision Making

Humans are limited both in their memory capacity as well as in their ability to consider multiple things simultaneously. The modality, the quantity, and the complexity of information all affect individuals' abilities to form decisions. Individuals with expertise in a decision making area may express these issues less since such individuals have existing heuristics and well-developed mental models (schema) to aid them in simplifying complex decision making tasks. Load can occur within (a) short-term memory capacity, (b) working memory capacity, (c) attentional resources, and (d) human perceptual capabilities. A tool will cognitively facilitate individuals' decision making process if it reduces the load in one or more of these four areas.

Humans have documented limitations in their short-term memory capacity. Short – term memory is thought to be limited to 4 +/-1 concurrent items (Parker, 2012). However, mnemonic devices and heuristic shortcuts can augment this relatively static limitation (Baddely, 2002). One example would be chunking where a large amount of information is divided into a smaller number of pieces, or 'chunks', of information. For instance, in order to memorize 10-digit phone numbers, learners will divide each number into three groups so that they are easier to memorize.

Working memory is similar to short-term memory except it relates to the amount of information that can be 'worked on' at any given moment. If short-term memory is the number of items on the table at a given moment, working memory represents the items held in your hands. Working memory is also limited and it varies in capacity across individuals as measured by such tasks as span tasks, anti-saccade tasks, the Stroop task, and dichotic listening tasks (Chepenik, Cornew and Farah, 2007; Baddely, 2002; Baddely & Hitch, 1974; Engle, 2002). For the purposes of cognitive tools such as the one related to this study, the working memory capacity appears to be related to cognitive load and efforts to reduce cognitive load through increasing the efficiency of information should increase the working memory capacity for users.

In addition to working memory storage limitations, working memory is also prone to storage and retrieval difficulties. Proactive interference refers to the effect that previous information has upon current information processing (Engle). This type of interference can make it difficult for individuals to remember what they were about to do or to remember the next step in a sequence of events. A more efficient cognitive tool should serve to reduce proactive interference through organizing information according to the task. In addition, a cognitive tool that prevents the build-up of proactive interference should augment working memory.

Attention, as a resource, is also necessarily limited both in scope and in size. The amount of information that an individual can consider at any one time is finite, is subject to individual differences, and decreases over time (Driver, 2001; Fernandez-Duque and Johnson, 2002). Although there is not a strong consensus on how attention works or whether it is even a phenomenon, there are ways to facilitate time-consuming mental processes. Some treat attention as a central executive that makes internal decisions; others treat attention as a discriminating spot light that finds salient information (Baddely, 2002; Wolfe, 2003). Information that is more salient, more personally meaningful and that involves multiple senses has an increased likelihood of rising above the ‘limen’ required for the phenomenon of ‘attention’ to occur (Greenwald, 1992, Fernandez-Duque and Johnson; Wolfe; Baddely). The cognitive tool can work towards facilitating limited attention resources through presenting information in an efficient manner. To increase the efficiency of attentional resources using cognitive tools create these tools so that they highlight information that stands out or that they include information that is personally meaningful for users.

Proper information visualization techniques should lead to a larger and more accurate amount of 'quantitative' evidence for the formation of curricular decisions. Vision is probably the most important perceptual modality. Humans initially establish their self-identity with vision (Wolfe, 2002; Tulving, 2002; Gallup, 1970; Rochat, 2003). We also know that humans learn about new information through a visual search pattern that emphasizes new information (Wolfe, 2002; Rensick, et al., 1997). Of course, the type of visualization should consider the natural human inclinations as well. For instance, a good visualization will account for both the ‘gestalt’ like perceptual tendencies as well as the detail-orientated nature of cognition (Kosslyn, Thompson and Ganis, 2002; Pylshyn, 2003). Although vision is important, using multiple modalities, such as sound-enriched visualizations, to augment human perceptual abilities may further increase the level of detail as well as the accuracy of the resultant categorizations (Sukhoy and Sinapov, 2010). The development of the dashboard system will focus primarily upon visualization of information. Multiple modalities would be areas for future work.

Liu et al (2004) classified nearly half of the 13 tools in their study as tools that reduce cognitive overload. All of these tools share the ability display information related to concepts of interest. For example, the research room tool was a tool that just contained information about various aspects of physiology, technology, and history. The periodic table tool helped students analyze the data that they collected in the context of the domain of knowledge they were studying. These types of tools help people make better decisions because they show information in a nature congruent with natural human analytical and visual abilities (Watson, 2004). This dashboard will attempt to make curricular information more organized, more accessible, and more visually optimal.

Scaffolding-Based Approach to Facilitating Decision Making

Vygotsky (2010) proposed the notion of social scaffolding, the idea that society provides the necessary infrastructure for complex human thought (see also Salomon 1988). Liu, et al (2004) discuss how cognitive tools act as a scaffold when they help individuals make decisions they could not normally have made without the assistance of a cognitive tool. Cognitive tools should facilitate these ‘out of reach’ decisions in addition to augmenting an individual’s normal problem solving ability. Jonassen (2003) mentioned how cognitive tools and similar support systems externalize decision makers’ internal representations of problems. This process varies based upon the complexity of the problem, but relates to transferability of a problem-solving style to new situations. From this perspective, a cognitive tool should be able to assist decision makers through increasing transference rates of existing problem solving abilities to new situations or through assisting with the structuring of complex problems to make them easier to solve.

Vygotsky characterized social scaffolding in terms of the *Zone of Proximal Development* (ZPD). In the context of cognitive tools, the ZPD relates to individuals’ ability to make decisions with cognitive tools that they could not have made without them. While the zone of proximal development relates to progress an individual makes during the use of a technology, internalization relates to development that occurs because the user had previously used a technology (Salomon, 1981). In order to think in terms of a tool, a user needs to internalize a tool. A user internalizes a cognitive tool when the use of that tool helps them categorize information and formulate problems on their own. Compare cognitive tools by determining which contributes to better user internalization. Based upon this definition of internalization, the dashboard will act as a cognitive tool if it increases individuals' level of structure related to the domain of knowledge associated with making curricular decisions.

Liu et al (2004) offers some guidance as to how to construct a tool that facilitates out-of-reach activities. The probe builder room is a tool that helped students design space vessels through giving them examples of past vessels that met certain requirements. In the same study, the launcher room tool similarly helped students design vessels based upon budget constraints. These tools both provided examples yet also offered constraints.

The dashboard will provide scaffolding in two areas. First, it will create a common language related to the curriculum. Language acts as a cognitive tool when it facilitates thinking about concepts (Mirolli and Parisi, 2009). Second, the dashboard should increase a user's level of structure in the curricular knowledge domain. While a user is forming a decision, the language used within a cognitive tool should increase both the number of concepts as well as the comparisons between those concepts that a user considers. Note, however, that curricular decision making experts might not need social scaffolding as much as non-experts do because experts already have a richly established hierarchy of related knowledge.

Attribution-Based Approach to Facilitating Decision Making

Augmenting an individual's ability to form and test hypotheses will also facilitate their decision making ability. This idea is similar to 'evaluating process and outcome' mentioned earlier, however, it is distinct in that here it is about facilitating individuals' abilities to make categorizations that are more accurate and to form judgments that are more rational. Here, cognitive tools facilitate decision making through augmenting humans' natural abilities to form attributions of causality and to categorize events.

The ability to form accurate categorizations and rational judgments is important because individuals tend to detect trends that fall into pre-existing thought patterns (attributions) about reality. This tendency can lead to perceptual and categorical errors such as the tendency to resist forming new concepts (as illustrated by prototype theory), make unnecessary generalizations, form inaccurate stereotypes, and exhibit escalation of commitment. The tendency to resist forming new concepts occurs when individuals tend to accommodate existing information into their current set of concepts rather than creating a new concept. Authors who discuss prototype theory detail how the formation of mental categories tends to effect the categorization of future information (Holyoak, 2008; Medin, 1989). Generalization occurs when individuals add details to new perceptions using data from previous, similar perceptions (Crain, 1985). The new information that does not relate to existing categorical structures is usually marginalized (Collins and Quillian, 1969; Medin, 1989; Crain, 1985). Similarly, stereotyping occurs when an individual overestimates the likelihood that an observation falls into a category based upon an incomplete set of features (Gray, 2004; Collins and Loftus, 1975; see also Medin). Finally, the escalation of commitment relates to individuals' tendency to favor previous information and previous decisions (Drummond, 1995).

In addition to perceptual and categorical errors, there are also documented cases of biases in human decision making. When it comes to making rational decisions, Tversky and Kahneman (1974) have documented numerous innate biases that humans suffer from including: representativeness of information, probabilistic reasoning errors, and problems with anchoring perception. People also subconsciously form biases in their impression formations when the categorization task invokes their self-identity (Craik and Tulving, 1975). Similarly, the fundamental attribution error refers to the tendency for individuals to attribute the etiology of favorable personal events to an internal locus of control while attributing the etiology of favorable events within other's situations to external loci of control (Jones and Davis, 1965). This is not to say individuals purposefully distort their decisions; it is simply human nature. A cognitive tool needs to present information in an objective and consistent way in order to facilitate accurate and unbiased decision-maker attributions.

The errors and biases found in human decision making occur because of our natural ability to create cognitive shortcuts unconsciously. Although sometimes harmful, if properly facilitated, these heuristics can aid individuals in making decisions. Heuristics simplify complex patterns into simpler rules; categorization can also aid with decision making since large elements of data can be 'chunked' into categories. Later, the recollection of these categories serves as an aid, or a tool, for the recollection of the original event. The recent popularity of memory gyms illustrates how heuristics and similar techniques work. Memory gyms, used for competitive memorization, train people to recall information through a variety of techniques such as heuristics, chunking, mnemonics, or other forms of aided recall (Foer, 2011). Incorporate such mental shortcuts into a cognitive tool through focusing upon the language of and the organization of concepts used to display information. The dashboard related to this project utilizes a common, concisely organized language and places it into a hierarchical structure based upon the relatedness of concepts.

Liu et al (2004) give two examples of tools to help students form and test hypotheses. The control room tool helps students decide on a course of action based upon the results of their previous investigations. The solution form tool encourages students to record their rationale for each decision. The combination of unified display of historic research alongside a rationalization present inside these two tools helps students to reflect on the decisions they made. This helps students to gauge whether their decisions were properly informed.

When considering an attribution-based approach to facilitating decision making, data should be openly available and transparent. In instances where information is guarded, expensive, or unavailable, decision making becomes more of a herding behavior where individuals' decision making processes tend to follow the decisions made by a group (Yalamova, 2009). This tendency to follow the crowd is similar to groupthink, a concept where the decisions made by a group can be different (often poorer) than the decision that any one individual would have made (Janis, 1971). Although it is problematic to try to measure the complexity and uncertainty of information, the dashboard can help mitigate problems such as groupthink because when data are openly available, decision makers are more likely to use data to support their decisions rather than simply following the crowd.

Process-Based Approach to Facilitating Decision Making

The final proposed cognitive function through which cognitive tools facilitate human decision making is the cognitive process itself. Here the cognitive tool works by augmenting the individual's meta-problem solving ability. Facilitate this by facilitating the other three cognitive tool functions. This function has four separate cognitive processes: *understanding the problem*, *identifying, gathering, and organizing information*, *integrating information*, and *evaluating process and outcome*. These processes come from Liu, et al (2004) who all made an effort to remain congruent with Bloom's taxonomy and the IDEAL problem-solver (Brandsford & Stein, 1985, as cited by Liu, et al, 2004).

As mentioned earlier, a study done by Liu, Bera, Corliss, Svinicki and Beth (2004) provides the categorizations of cognitive processes that I will use to discuss problem solving. This study examined 21 components of a set of cognitive tools that teachers created to help sixth graders solve different types of problems. Experts categorized each cognitive tool component based upon its likely cognitive function. They also categorized different types of problems based upon which cognitive processes were required to solve them. Liu, et al then looked at usage patterns of all 140 students and examined how the frequency of cognitive tool use related to each cognitive process. They determined which cognitive tool functions participants used more or less frequently for each of the cognitive processes used during problem solving.

Understanding the problem

Liu, et al (2004) discovered that students reported using tools with a cognitive-based function in order to understand problems significantly more than was expected. Students reported using scaffolding-based and attribution-based cognitive tools significantly less than expected when it came to understanding problems. According to these results, in order to aid in the *understanding the problem* process, a cognitive tool should work to reduce cognitive load.

Identifying, gathering and organizing information

When it came to the *identifying, gathering, and organizing* process, Liu, et al (2004) found that students used scaffolding-based cognitive tools significantly more frequently than the other types of cognitive tools. In contrast, students used cognitive-based tools significantly less than expected for identifying, organizing, and gathering information. They concluded that the *identification, gathering, and organization* process of information required cognitive tools that could help users perform otherwise out of reach activities.

Integrating information

Students used both the attribution-based and cognitive-based tools significantly more during the *integrating information* process. Conversely, students used scaffolding-based cognitive tools significantly less frequently than expected. The authors concluded that the *integration information* process likely required tools to support cognitive load and hypothesis testing.

Evaluating process and outcome

Students used cognitive-based cognitive tools significantly more frequently than expected in the *evaluation of process and outcome* process. Conversely, students used scaffolding-based cognitive tools significantly less than expected in the *evaluation of process and outcome* process. According to these results, this process likely requires tools to support cognitive load.

Other considerations

In addition to cognitive tool function, the difficulty level of the problem will also likely interact with these cognitive processes (Jonassen, 2003, Liu et al). Jonassen (2003) discussed how three different types of cognitive tools help to solve problems: semantic networks, expert systems, and systems modeling tools. Jonassen wrote that tools help individuals solve problems by helping users to represent problems through externalizing their internal representations. Based upon Jonassen's research, it seems as though cognitive tools could fully externalize easy problems. Harder problems are harder to externalize. Based upon

Jonassens' research, easier problems should utilize more of the first two processes: *understanding the problem* and the *identifying, gathering, and organizing information* while harder problems will use more of the later processes: *integrating information* and *evaluating process and outcome*.

Some similar elements are present across all of the elements of process-based cognitive tools. Liu et al (2004) offer three examples of such tools. The notebook allows students to record information but it forces them to do so without the aid of copy-and-paste tools. The bookmark tool allows students to drag images that they find during their research into a single portfolio for later use. The expert-modeling tool provides a similar capability for videos. These process-based tools help students keep track of what they find interesting as they explore the decision making process. Through using these tools, students are gathering and integrating information as they process which decisions to make.

In summary, Liu et al (2004) found that there is an interaction between tool usage patterns, problem types and cognitive process. People use the elements of a cognitive tool with varying frequencies depending upon whether they are using the tool to solve a simple or a complex problem. This means that it is possible to infer which parts of a cognitive tool people used during each of the four cognitive processes associated with decision making. Doing this requires being able to identify what types of tools are useful during each cognitive process combined with the knowledge of whether usage frequencies should be higher while solving easy or while solving hard problems. While the study by Liu and Bera provided insight regarding the relationships between cognitive tools and cognitive functions, it did not experimentally explore how specific cognitive tool designs for addressing these cognitive functions can account for differences in user performance. One goal of this project is to address this gap in the literature.

The Context

The context for this study is two-fold. First, the interest and motivation for this study have come from a theoretical context. Second, there is a practical need for a system that can facilitate curricular decision making for faculty and other decision makers at the ISUCVM. The ISUCVM has recently developed a cognitive tool for making data-driven decisions about curricular change. This tool will play a role in curricular assessment activities of the ISUCVM , including examining student outcomes and making recommendations for change (Fitzpatrick, et al., 2004). The motivation for the use of the dashboard to help with curricular decision making comes from its ability to represent the complete curricular context concurrently, to facilitate decision making through presenting important supporting evidence at the right time and place, and to facilitate collaborative decision making through tying student outcomes to common taxonomy.

Curricular decision making requires understanding the complete context surrounding a particular curricular decision. For example, one technique for curricular evaluation is teacher-based assessment. This form of assessment examines the performance of professors as evaluated by students. This form of assessment is tricky to use for curricular decision making because it contains at least two dimensions: 'student performance' and 'student ability to handle academic pressures' (Bowers, 2009). The multiple dimensions within teacher-based assessment only became evident to researchers when they looked at their data holistically. Holistic analysis has become more common; a summative assessment process that considers the complex interactions within a curricular environment warrants a holistic analysis (Patton, 2002 p. 54-61). A dashboard interface permits a holistic approach to understanding the curricular system through allowing decision makers to compare outcomes data for the entire curricular system concurrently.

The evaluation of the use of the dashboard should focus upon the ability of decision makers to make higher quality decisions. Increasingly, the goal will be to move away from simple recollection, understanding, and application towards analysis, evaluation and the creation of insights that are useful for making decisions (Bloom, 1971). According to Bloom, this is the best way for formative assessment to support the distillation of knowledge. Similar to this concept from Bloom is the analytical induction framework (Worthen, Sanders & Fitzpatrick, 1997) which emphasizes the ability to triangulate on a decision with evidence from multiple sources. When making curricular decisions, there is a difference between data and evidence (Knapp, Copland, and Swinnerton, 2007). It is more important to focus both upon the users' final decision and upon the evidence that users reported using than it is to focus upon the details about what users were looking at or doing.

Cognitive tool developers need to understand which concepts are involved in solving problems in the target domain, and how those concepts relate to each other. For instance, concepts involved in medical problem solving include body systems and processes, which the expert physician must understand in order to solve medical problems. Supporting evidence-based decision making with a cognitive tool first requires the creation of a common language. In the context of creating a cognitive tool for curricular problem solving in a specific setting, the relevant concepts include courses, competencies, stakeholder satisfaction, and student performance across veterinary content areas. The ISUCVM created this representation of concepts using independent accrediting agency standards, and results from student, alumni, and employer surveys. Creating this clearly articulated set of concepts was important for data aggregation and for efficient collaboration.

The practical context is useful to test the ideas represented by the theoretical context discussed in detail during the introductory section of this paper. Furthermore, several key connections between the practical and theoretical context make it possible to test whether the dashboard facilitates decision making in this given context. Finally, the practical context offers an opportunity to address several gaps between theoretic and applied contexts within the literature.

Hypotheses

Domain experts have more highly developed internal schemas that make them naturally more capable of simplifying information related to solving a problem within that domain. Technology experts will have more knowledge about how interfaces function and will likely exhibit more efficient usage patterns. In both cases, the amount of evidence used by the expert will differ from that of a novice. In the case of the domain expert, fewer concepts should be required to form good decisions. The technology expert should exhibit fewer exploratory behaviors. This study examined both of these hypotheses using the concept tracking and mouse-tracking data associated with each participant.

Main hypotheses

The primary hypothesis for this study was that the structure and function of each region of the dashboard would affect usage patterns between the five regions. To better explore the questions presented during the literature review I then broke this hypothesis into several related questions.

Recall that the user interaction within the dashboard will vary across regions because certain regions are not accessible without explicit prior use of other regions. There is a three-stage model that affects the overall likelihood for concept consideration and mouse behaviors to occur within each region. This stage effect results from how region one is always accessible but regions two and three are only accessible if the user clicks submit and enters into the second stage of the dashboard. Similarly, regions four and five are only accessible if the user chooses to examine what they find in regions two and three in more detail. For this reason, I also hypothesized that the later stages would be used less often.

H0: Participants will use region 1 more than regions two and three; participants will use regions two and three more than regions four and five.

Based upon the work of Liu and Bera (2005), participants should use tools that support scaffolding and attribution more than tools that support cognitive processes and cognitive loads. Participants use these process-based tools more for solution generation while they use cognitive-based tools more for research related to problem solving. Jonassen talks about experts naturally using information more efficiently and this expresses itself as a reduction of irrelevant evidence use. As this relates to the use of the dashboard, there should be a reduction in both the types of behaviors exhibited and the frequency of occurrence of each of these types. Since regions one and two are both cognitive-based regions, participants will use them more for research-related processes while they will use process-based regions four and five more for solution generation.

H1.1: The higher performing group will have a lower overall number of exploratory behaviors and of number of concepts considered.

Individuals who make better decisions using the dashboard will likely utilize more of its solution generating potential. Similarly, the impact of the stage effect will be smaller and these participants will exhibit less regional variation in usage patterns.

H1.2.1: The proportion of use for regions four and five will be significantly higher within the higher performing group

H1.2.2: The proportion of use for regions one and two will be significantly lower within the higher performing group

Covariate hypotheses

In addition to differences in usage patterns between high versus low performing individuals, there are also several covariates to consider. As mentioned, individuals tend towards technology that suits their strengths. Hypothesis 2.1 discusses how individuals with higher spatial acuity will likely gravitate towards the visual parts of the interface.

H2.1.1: The proportion of use for regions two and three will be significantly higher for individuals with higher spatial acuity.

H2.1.2: The proportion of use for regions one, four, and five will be significantly lower for individuals with higher spatial acuity.

As previously discussed, the adoption of technology by a user relates to their previous experience with similar technologies. More experienced individuals will prefer technologies with which they are familiar. These individuals have previously learned efficient behaviors from exploring similar technologies. These users will use fewer behaviors to accomplish their goals since they can build upon this experience. Hypothesis 2.2 discusses the expectation that individuals with HI technological expertise will exhibit higher overall performance. This increase in performance may cause differences in regional variation within usage patterns that should be more similar to those hypothesized for the HI performing group. Hypothesis 2.3 discusses these expectations.

H2.2: Individuals in the HI performing group will have significantly higher overall scores on the experience with technology survey.

H2.3.1: The proportion of use for regions four and five will be significantly higher for individuals who score HI on the experience with technology survey.

H2.3.2: The proportion of use for regions one and two will be significantly lower for individuals who score HI on the experience with technology survey.

CHAPTER 2

METHODS

Materials

The goal of the newly built dashboard system is to assist decision making in the previously described context. This materials section focuses on the various structural components of the dashboard. The rest of the sections in the methods discuss the design of the experiment.

There are five major regions within the ISUCVM dashboard-styled interface. These regions are: 1) Tree View, 2) Table Graphs, 3) Annotated Timeline, 4) Question View, and 5) Metadata View (see Figure 1). It is possible that users who perform well will use certain pieces of the cognitive tools more or less frequently than those who perform poorly. This is because each of these regions will likely differentially facilitate each of the four functional approaches of cognitive tools. To understand why this could be the case, first consider the function of each section separately.

The first section called the ‘tree view’ represents a conceptual hierarchy of terms related to the context: curricular development in veterinary medicine. There are three levels within this hierarchy. The highest level represents broad concepts such as a student’s clinical competency or their assessed skills in basic science. This ontology developed from a combination of external standards set by the American Veterinary Medical Association, historical practices used within the college, or best practices recommended by decision making experts. Selecting the root of the hierarchy will implicitly select the related child concepts. However, it is possible to expand the tree and specify concepts. An individual could interact with multiple concepts within this section. When they click submit, their choices are then persisted to other regions within the dashboard.

Section two contains tables of graphs that show descriptive level statistics related to each term specified by the user in section one. This section shows a multitude of information for each concept. Starting from left this section first shows the names of the concepts. Next, it reports the historical average as a numerical value followed closely by a longitudinal illustrating historical performance. The values for all concepts were so that their displayed values range from one to five where five is the maximum possible score. This section then reports current performance next to a bullet chart. The bullet chart compares the current overall performance to the performance goal for that particular concept. This curricular context uses a criterion-based standard for all concepts. Here the criterion for success is obtaining a score of at least three out of five. Next, this section shows the distribution of scores for each concept. The minimum, first quartile, second quartile, and maximum values appear immediately to the left of a quartile plot illustrating these same context in graph form. Finally this section illustrates the text and graphical version of confidence related to each value. This value is a percentage that ranges from zero to one where one represents a hundred percent confidence in a score. An individual can sort concepts based upon the average, current performance, distribution, or confidence values by clicking on these related headers. They can also select one or more of these concepts to examine in further detail.

The third section primarily contains an annotated timeline, hence the name of this section “Annotated Timeline.” A request for details in either the first or the second section will cause the dashboard to graph these concepts over time across this timeline. The user can select the area of interest by interactively changing the window on the bottom of the graph.

Metadata related to events for each concept show up as letters upon the timeline. To the right of the graph these same letters appear alongside a more detailed description of these events. This whole section is interactive in that moving the cursor along the timeline will illustrate the numeric value associated with the location of cursor point.

After a user has interacted with the second section, any selection the user chooses to see in more detail will appear in the fourth and fifth sections. The fourth section shows all of the questions presently associated with each concept. Organized as a tree view, this section appears similar to the first section. This section used same ontology used by the first section. Within each concept, any related questions appear alongside the original score used to assess these measures. The fifth section describes a sequence of events related to student outcomes and to the curriculum. For instance, consider a cancelled course. In this case, the affected concept along with the cancel date appears with the description summarizing the course cancellation.

At any given time, the system is aware of which concepts are present in each of the five regions. The cognitive tool generates information about Veterinary Medical education concepts on the fly in three different stages. In the first stage, the tree view (region one) always contains a list of all concepts. When individuals select items and then click ‘submit’, the system graphs these items. Upon clicking submit, the system enters the second stage. Here the dashboard lists the selected concepts in both the Table Graphs (region two) and the annotated timeline (region three). To move to stage three, participants interact with a piece of the region two that allows them to ‘examine selected constructs’ in more detail. Upon clicking this button, users enter the third and final stage. This stage displays information related to selected concepts in the remaining regions three, four, and five.

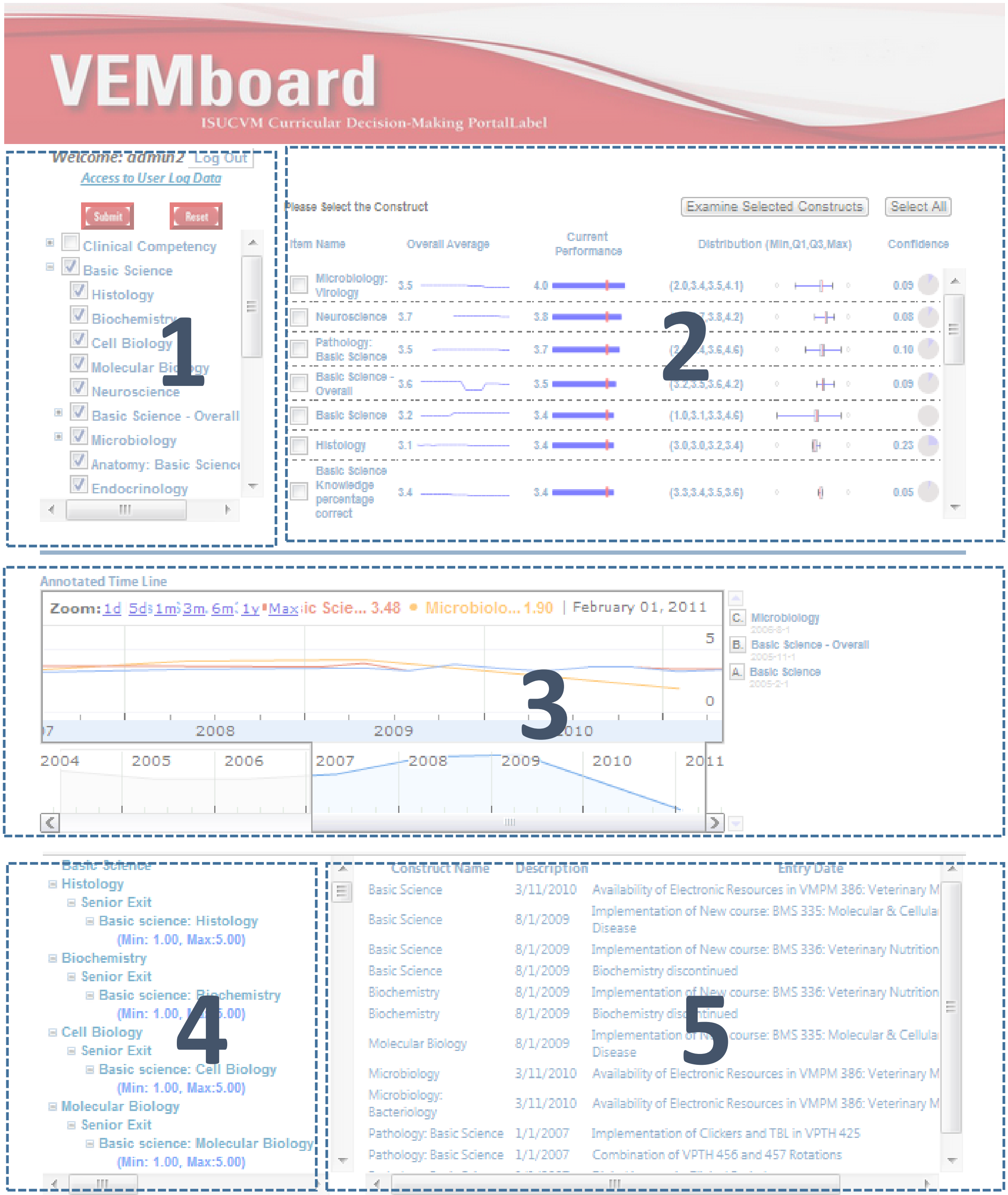

Fig. 1: The ISUCVM Dashboard interface and its five major regions: 1) Tree View, 2) Table Graphs, 3) Annotated Timeline, 4) Question View, and 5) Metadata View.

As mentioned the usefulness of each section for decision making should vary. This is because I designed each section to facilitate a slightly different cognitive function. Figure 2 shows the results of categorizing each region according to the cognitive function(s) it is designed to facilitate, using the definitions provided by Lajoie (1993). This taxonomy has been used by several authors (Liu and Bera, 2005; see also Liu, Bera, Corliss, Svinicki, and Beth, 2004; see also Kim and Reeves, 2007). This diagram maps the four cognitive functions to one or more regions. Regions one and two are primarily cognitive-based since they support both cognitive load and cognitive process cognitive tool functions. Regions four and five are primarily process-based since they support both scaffolding and attributional cognitive functions. Region three could be either cognitive or process-based.

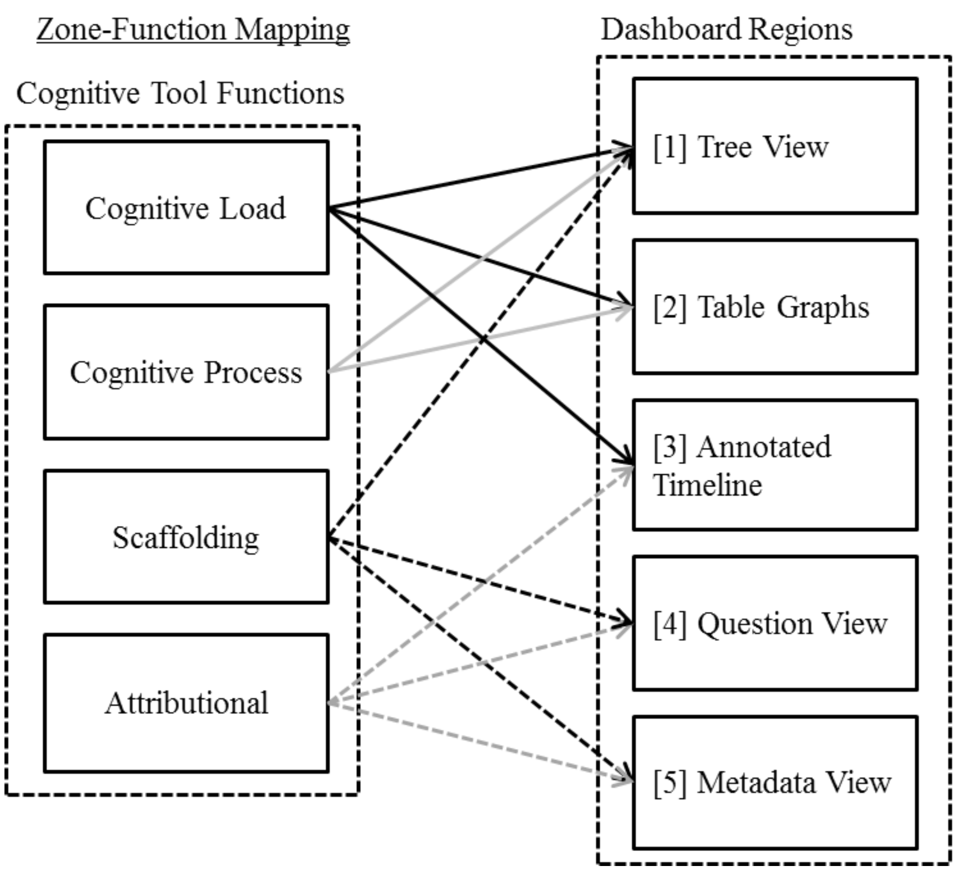

Fig. 2: This is the theorized visual mapping between cognitive function and region.

Through looking at the behavioral patterns exhibited within each of these five regions, and through considering the work of Liu and Bera (2005), this study will ultimately classify each of these five components based upon outcomes data. Through comparing the results of better performing participants to those of worse performing participants, I should know which behavioral differences, in terms of each region, might be accounting for performance differences. The experiment outlined in this study helped to test these ideas.

Procedures

The purpose of this study was to learn whether people who are better at solving problems use each piece of the cognitive tool differently. Recall that Lajoie (1993) proposed four major approaches cognitive tools could use to facilitate problem solving: cognitive-based, attributional-based, scaffolding-based, and process-based approaches. Liu and Bera (2005) examined these four approaches and found that the frequency of use of a tool using each approach differed depending upon the type of problem. Rather than altering the type of problem, this study used several very similar problem sets to examine how individuals use each piece of the interface. This study then combined this knowledge with information about each participant to examine whether these differences co-vary with individual differences. The following describes an IRB approved study that I conducted in order to answer these questions.

Since the goal of the study is to measure differences in usage patterns between interface regions, the study needed several simulated situations so participants could use the dashboard to make informed curricular decisions. Theses simulations took the form of three problem sets. I interviewed several curricular decision makers to ensure that each problem in these sets had a roughly uniform difficulty. The instruments section discusses these tasks in more detail. This study also counter-balanced the order in which each participant received each of the problems. All participants answered the same questions but the order of the questions differed depending upon to which group the random selection assigned each participant.

In addition to controlling the participant tasks, the testing environment was highly controlled. I led participants through the informed consent and through the spatial acuity test. All participants used the same exact computer and browser combination. I offered a brief overview of the dashboard, showed participants how to login, and answered questions. Then participants had three minutes to explore the dashboard so they could see how the interface worked. During this time, participants were also free to explore the content domain. After the exploration, participants started their tasks. I instructed them to try to spend no more than fifteen minutes on each task. After about 45 minutes, most participants had finished their tasks. Two participants took longer than expected but all participants finished within an hour. Even after the last participant was done with their tasks, the design of the experiment was just beginning. This study addresses whether there are differences in usage patterns across the five components of a dashboard between individuals who are better or worse at solving problems. For this reason, this study gathered data about usage patterns that may contribute to these differences in performance. These data came from various sources such as the types and numbers of concepts considered or of behaviors exhibited. Through combining evidence from a variety of sources, it will be easier to analyze which elements of the interface were more or less helpful for participants.

Exploratory behaviors are useful for understanding how individuals are using an interface to externalize a problem. These behaviors relate to developmental problem solving

(Sukhoy and Stoytchev, 2010). Since eye tracking and mouse tracking record natural and unconstrained interaction between a user and an interface, these two techniques can record participants' overt, exploratory behaviors. Eye tracking is similar to mouse tracking and both techniques collect natural human interaction patterns useful for examining differences in behavior. Techniques such as these have already detected differences in evidence use patterns between novices and experts (Sohn, Douglass, Chen and Anderson, 2005). After examining 595 web pages, Chen, Anderson and Sohn (2001) found that mouse cursor and eye fixation location were directed toward the same region more than 75% of the time. These authors also found that the pixel distance between eye fixation point and mouse cursor location was relatively uniform and that when both mouse and eye saccades are to a meaningful region, the distance between the two is minimal. Finally, mouse tracking is a more naturalistic method than eye tracking because it does not require users to have any special training or apparatuses. For these reasons, this study used mouse tracking to record participant interaction behaviors.

A tracking apparatus recorded a multitude of user behaviors including time, distance, velocity, the number of concepts considered, and the number of mouse movements between regions. The instrument section describes this in more detail. All of these behaviors were dependent variables in this study. The number of mouse movements between regions refers to how often a users' mouse moved between the five regions. Time refers to the average time between mouse clicks. For instance, a time of 1 second would record a participant mouse click at 12:00:00 and a click at 12:00:01. Similarly, distance measures the number of pixels between each mouse click. Velocity is the quotient of distance divided by time for each click. Recall that the cognitive tool has three stages of interaction and that stages two and three allow for interaction that is more detailed. The number of concepts considered referred to the total number of items a participant decided to examine in detail.

An expert panel consisting of three faculty from the ISUCVM helped to evaluate the quality and accuracy of participant responses. This evaluation used a rubric created specifically to assess the quality of solutions across multiple dimensions. The instruments section describes this rubric. The consistency of the panel was important. In assessing agreement using Cronbach's alpha typically this score should be greater than or equal to 0.70 (Stemler, 2004) (see also Pedhazur, 1991). This study used these criteria to ensure that the assessments of performance by the expert panel met both of these criteria. In order to reduce the overall burden on each rater a single rater graded every participant and the remaining raters together each graded half the participants. This resulted in two Cronbach's alpha values 0.875 and 0.711 for each couplet as well as an overall value of 0.833 for all raters.

The study examined the performance of individuals as they completed three sets of problems using the newly create dashboard instrument. All of the problems simulated real curricular decision making contexts. The problems were as similar to each other as possible and yet the order of presentation to participants was still counter-balanced. The experimental environment was highly controlled. The mouse tracking and the concept tracking apparatuses custom built into the dashboard kept track of participant behaviors that took place during the experiment. After the participants completed their tasks, the experiment continued as experts rated the performance of each participant. The final analysis for this experiment examined the differences between user behaviors

Participants

Jonassen (2003), Vygotsky and Salomon suggest that expertise could play a confounding role in problem solving. Jonassen (2003) suggests experts should require less information to make decisions due to their more highly developed internal schema. When it comes to cognitive tools this makes sense. According to Jonassen, one of the major functions of a cognitive tool is to externalize information that previously required internalization. This externalization should reduce cognitive load. Salomon mentions that cognitive tools can shape thought patterns. This sort of 'AI in reverse' suggests that those who have more richly developed internal schemas may be less satisfied with a tool that is incongruent with previous experience. This relates to Vygotsky's notion of the 'zone of proximal development' in that experienced individuals may not benefit as much by scaffolding as less experienced individuals would. More experienced users may not use cognitive tools that function through scaffolding. Furthermore, such individuals may not benefit from these types of tools. In order to account for these differences in behaviors, the present study split participants based upon their ability to solve problems using a cognitive tool. Liu and Bera (2005) split participants based upon performance using cluster analysis and then examined differences in behavioral patterns between these two groups.

The present study used similar clustering analysis to separate participants based upon their solution accuracy into a 'better' and a 'worse' group of problem solvers (Xu, 2007). This study split participants into several types of groups. The first type of group existed solely for the purposes of counter balancing the order in which participants received questions. Three additional groups split participants based upon their performance, their experience, or their spatial acuity. The analysis section further discusses these participant classification techniques.

This study received IRB approval and all participants were free to opt out at any time during the study. Sample size was determined by examining the results of a pilot study that I discuss in the next section. Based on this study, power (>=0.80) to detect significance should occur using this design as long as 3-4 people participate. In total 12 people participated in this study. Each user completed three trials. There were two levels of group and five levels of interface to consider. The total number of observations for each participant (N) was (3*5)N/(2*5). Thus, given the design of this experiment, the size of the data grew by 1.5N for each new participant.

In line with previous work and with the context of this study, I recruited participants from a panel of individuals familiar with making curricular decisions. At the ISUCVM, this group of people consists of professors, members of the curricular committee, and members of the collegiate administration. Given the relatively small group of individuals from which I could recruit and given the admittedly low response rate, it would have been unreasonable to use random selection on the group of participants who expressed interest. Instead, I used random assignment to randomly place individuals into groups.

Pilot study

The effect size estimates from a preliminary pilot study helped guide the estimated number of participants needed for this study. Invitations to participate in this pilot study went out to the same cohort of curricular decision makers that I later invited to participate in the primary study. Three volunteers participated.

There were a few differences between the pilot study and the final study. While the pilot study was interested in comparing the dashboard interface against an existing control, the final study was interested in examining the decision making capabilities of each region within the newly created dashboard interface. The pilot study also did not examine spatial acuity or participant experience. A more subtle difference between the two studies was the number of user behaviors recorded. The pilot study just measured the number of concepts considered as measure of user behavior. Nonetheless, participants still examined various concepts in conjunction with the same questions that participants in the final study later completed. Also similar to the final study, participants in the pilot study were still asked to rate their levels of satisfaction with the interface.

The pilot study yielded both a participant-level and an interface-level effect size estimate. I used both of these estimates to determine the minimum number of participants to recruit and to determine the number and types of tasks that each participant would need to complete based upon a desire to have at least 80% power with a 95% confidence. The participant-level effect size estimate came from examining the differences in satisfaction levels across participants. This estimate had a mean of 4.74 with a standard deviation of 1.98. Based upon this estimate I would need to recruit at least 91 degrees of freedom in order to determine significant differences in interest using a within-groups design. Since each task largely focused upon a different region of the dashboard interface, the interface-level effect size estimates came from examining the variances in user behavior across different tasks. This estimate had a mean of 266.66 and a standard deviation of 276.83. Based upon this estimate I would need a combination of at least 17 participants and tasks to detect significant differences in user behavioral patterns using a within-groups design. Clearly, the interface-level design will be much more powerful given the recruitment constraints.

Sometimes the subjective insights gained from running a pilot study help to influence the design of the final experiment. This is also the case in this present study. Following the pilot study, the final study design changed to a comparison across the five primary regions within the dashboard interface. Since the focus of this research is on facilitating decision making using cognitive tools, the primary reason for changing the study design was to align more closely with this research. However, it was also hard to manage an introduction to two unique interfaces with users during a short time period. Finally, the effect size estimates offered from this pilot study combined with practical recruitment restrictions led to the final decision to focus primarily upon a within group comparison of the interface. Final recommendations based upon the pilot study focused upon boosting the effect size of the data through creating a more sophisticated user-tracking technology. This behind-the-scenes system would now track not only the concepts that users engaged in but also their behaviors they exhibited during their exploration of various dashboard regions. One of the practical and painstaking implications of this change was a movement towards client-side, java script based user tracking. This required a new system to integrate on top of the existing server-side user-tracking technologies.

Instruments

Several research instruments were necessary to ensure experimental efficiency or to test several direct and covariate factors. Where possible this study adopted instruments from existing apparatuses. At times, this study modeled instruments after the approach of other research.

The present study used the 'group embedded figures test' to measure spatial acuity. This variation of a paper-folding test examines individual differences in the perception of field dependence (Oltman, Raskin and Witkin, 1971). The embedded figures test only took about fifteen minutes to administer and was easy to include just prior to introducing participants to the interface.

The present study used an experience with technology survey (see Appendix) based upon the work of Maraka, Johnson, and Yi (1999). The goal of this survey is to categorize participant familiarity and experience with technology through recorded individuals' self-reported abilities with certain key technology functions.

Mouse tracking techniques were custom-built into the dashboard. This mouse-tracking system recorded where and when participants pressed a mouse button. This tracking actually provided several dimensions of participant behavioral data. First, this tracking provided aggregate behavioral data useful for understanding usage patterns between different regions. Similarly, this technique captured the delay between mouse events occurring within the same region. By recording when a mouse cursors enters or leaves a region, this tool was also able to track where a mouse cursor was upon a mouse button click. Since speed represnets distance dived by time and since this tracking technique recorded both the location and time of each click, this technique also then recorded the speed of each participant's interactions.

This study strived to ensure uniformity across tasks by following several guidelines. First, all tasks should be solvable using the cognitive tool in a similar manner. Next, an expert panel chose the three best problems from a set of all possible problems. This panel had a simple evaluative criteria for choosing problems: 1) the ability to be solved using the interface, 2) the presence of agreement by the panel as to what a solution looks like, 3) similarity in difficulty levels, 4) similarity in perceived problem-solving method.

An expert panel helped to grade the success of participant responses to each of six questions for each task. The expert panel consisted of three members of the college that each had experience making informed decisions about curricular change. The rubric rated each response from really good (5) to really poor (1). The rubric allows the evaluator to indicate the plausibility of, the comprehensiveness of, the optimality of, the supporting evidence for, and comprehensiveness of evidence supporting each participants' response. These five dimensions encompass different dimensions of a successful answer. The plausibility and optimality of the response relate to the goodness of fit of this response to the curricular domain. Supporting evidence and comprehensiveness are both characteristic of expert decisions. Not every question made sense in combination with the rubric. For example, grading the comprehensiveness and optimality of solutions only makes sense when evaluating a holistic answer. For this reason, I informed committee members to leave these parts of the rubric blank for both the problem area or success area questions. The rubric gave members of the expert panel several dimensions against which to evaluate each participant.

Analysis Methods

Splitting participants into groups based upon performance

When it comes to performance, there are factors other than accuracy to consider. Did participants perform well because they were better at solving problems or because they took more time to make sure their answer was correct? Did participants who completed tasks fast do so because they were better at solving problems or because they decided to skip a few tasks? In addition to accuracy, this study also considered speed as a variable of interest.

Mennecke, Crossland and Killingsworth (2000) found differences within performance speed that related to the level of participant expertise. Lankton, Speier and Wilson (2012) examined several internet tools used for decisional guidance on the ability to facilitate knowledge acquisition (KA). Their study found that performance related to speed and that there tended to be a 'skewness' that could occur where participants would make internal tradeoffs for speed or accuracy based upon task complexity. Speier, Valacich and Vessey (1999) found similar interaction between speed and performance related to the study of decision making performance. Furthermore, Jonassen suggests that expert participants will not just be better at solving problems, but that they should also naturally be faster at solving problems. Jonassen suggests that speed relates to task difficulty. For these reasons, clustering participants into groups based upon performance should consider the relationship between speed and accuracy.

Analysis of differences between groups

This study was a 2x5 design comparing higher versus lower performing groups against five levels of interface region. There were also several covariates to consider. These included spatial acuity, individual levels of experience, and several demographic 'control' measures of covariance. For these reasons, I used a MANCOVA to assess for overall significance across these factors. When the overall model was significant, I used a post-hoc pairwise ANOVA to examine significant relationships between each independent variable. There were two types of dependent outcomes: mouse tracking and concept tracking. Mouse tracking contains data about when a participants' mouse enters or leaves a region and when participants click the mouse button. I quantified mouse tracking as the number of mouse clicks per region, the time spent within a region, how often each region is a destination of a mouse saccade, the amount of mouse movement per region, and the frequency of use of each region. The concept-tracking dimension contains information about which stage of comparison a user is in and which concepts participants compared at any given time. I quantified concept tracking as the number of concepts participants used within a region, the number of different comparisons made within a region, and the delay between using different stages of the interface. For all comparisons, I assessed significance at the 0.05 level.

Analysis of covariance

If participants who make better decisions use each piece of the interface differently than participants who make worse decisions, then it makes sense to consider several possible covariant factors.

Satisfaction relates to individual preference towards a technology based upon their experience. Individuals will likely be more satisfied with a technology that allows them to solve a problem in a familiar way. Similarly, satisfaction may relate to self-efficacy. Task-technology fit (TTF) relates to the usefulness of a cognitive tool (or piece of a cognitive tool) for solving a problem. Jarupathirun and Zahedi (2007) found that the task-technology fit related to self-efficacy. An individual's belief in their ability to solve a problem with a given type of tool may affect their performance. This can be measured by examining participant satisfaction levels. If an individual both performs poorly and has low levels of satisfaction with the cognitive tool, then there was a poor task-technology fit for that individual. Jonassen (2003) claimed that there was little research on whether individual differences related to cognitive tool use and perhaps this is an area for the field to study further. Nonetheless, Lajoie (1993) claimed that a lot of the difference between two participants' performance could be due to individual differences. An important individual difference to consider for visual-spatial based tools is participants' spatial ability. Jarupathirun and Zahedi (2007) examined two dimensions of spatial ability along with self-efficacy in an effort to explain differences in performance between two spatial decision support systems (SDSSs). The present study found that neither dimension of spatial ability was useful for explaining TTF. However, other studies have found that spatial ability is important for explaining differences in performance between more and less experienced users of technology (Mennecke, Crossland and Killingsorth, 2000; Egan, 1988). There is reason to believe individual differences could account for variation in performance.

In addition to the factors already mentioned, this study also examined several covariate factors including gender, departmental affiliation, and self-reported levels of familiarity with several web and visualization technologies.

CHAPTER 3

ANALYSIS

I first split participants into three separate groups in order to examine experience, acuity, and performance covariates. Each group has two levels: high (HI) vs. low (LO). The assignment for each covariate was independent; participants assigned to the HI performance group may or may not be the same as the HI experience and HI acuity groups. Analysis used a means squared error (MSE) clustering technique to maximize the differences between groups while minimizing the differences within groups (Xu, 1997). For acuity, this method split yielded an average HI v. LO acuity score of 6.83 v. 3.00 on section I and 8.00 v. 3.50 on section II (Appendix table 3). The average performance scores for the HI performance group ranged from 3.18 to 3.70 while these same scores ranged from 2.52 to 2.80 for the LO performance group. The average experience survey scores across all survey questions for the HI group ranged from 3.67 to 4.57 compared to the LO group range 3.00 to 3.83. In no case did any of the participants in the LO group have scores that were higher than those in the HI group.

Several hypotheses involved examining effects across several regions. The simplest way to represent these regional groupings in the data was to encode three features that represented these three groups. These groupings were 1) regions 1 and 2; 2) regions 2 and 3; and 3) regions 4 and 5. If the behavior occurred in any region within one of these groupings, then I labelled the field for that grouping as 1 and otherwise as 0. These region groupings allowed for analysis of performance, experience, and acuity by region. For instance, Hypothesis 2.1.2 discusses analysis across regions 1, 4, and 5. In this case, use group 2 since those features labeled as 0 for group 2 will all be regions 1, 4, and 5.

The proportion of evidence referenced in several hypotheses refers to all of the dependent variables except for concepts considered. Where a hypothesis refers to exploratory behaviors, it refers to the frequency of mouse clicks and frequency of mouse events as well as to the overall number of concepts considered. It was a priority to test all hypothesis using as few statistical tests as possible. Due to the nature of the data and the difficulty in integrating data from various sources, analysis still required several MANOVA and MANCOVA depending upon the hypotheses at hand. The combination of two MANCOVA tests examined most every hypothesis.

CHAPTER 4

RESULTS

Main Hypotheses

The main hypothesis H0 dealt with exploratory behaviors across regions:

> H0: Participants will use region 1 more than regions two and three; participants will use regions two and three more than regions four and five.

A MANOVA examined whether time, distance, or velocity significantly varied across regions. In this model, region was a significant contributor to the overall model (see Appendix table 1). Subsequent factorial ANOVA analyses revealed that only distance traveled had a significant relationship with region (see Appendix table 1). While the overall duration between mouse clicks for regions 1 and 2 appeared higher than for the other three regions, this difference on its own was not significant. However, the distance between mouse clicks was significantly higher for region 2 (Appendix table 1). Since mouse movements between regions use a separate data source, a separate ANOVA examined these movements. There was a significant relationship between region and the time between mouse movements (Appendix table 1). Time between clicks was relatively low for regions 1, 4, and 5 but was roughly twice as large for regions 2 and 3.

A complimentary hypothesis H1.1 considered whether performance was a factor:

> H1.1: The higher performing group will have a lower overall number of exploratory behaviors and of number of concepts considered.

A MANOVA test examined whether performance related to time, distance, and velocity. Due to the way the online A/B data system tracked users, two separate ANOVA tests were required. One determined whether there was a similar relationship within mouse events and the other examined the number of concepts considered. The MANOVA test revealed that performance was a significant contributor to overall exploratory behaviors (Appendix table 2). The results of the subsequent ANOVA tests showed no significant relationship between total events considered or total mouse movements between regions (Appendix table 2). A subsequent factorial ANOVA examined time, distance, and velocity separately. This analysis revealed that both time and distance significantly related to performance. An analysis of means showed that the HI performance group had significantly longer durations between mouse clicks yet significantly less distance and velocity between these clicks (Appendix table 2).

The last main hypothesis was actually a set of two hypothesis considered the types of evidence used during the decision making process:

> H1.2.1: The proportion of use for regions four and five will be significantly higher within the higher performing group

and

> H1.2.2: The proportion of use for regions one and two will be significantly lower within the higher performing group

Recall that an overall MANCOVA examined total mouse clicks and the differences between regions and across several covariate factors simultaneously (Appendix table 3). A subsequent set of factorial ANOVA tests examined the differences in overall events across these same factors (Appendix table 3). Across the mouse click behavior, there were significant differences between regions 1,2 and regions 4,5. Within the events data neither of these two regional grouping was significant.

Covariate Hypotheses

In addition to examining the main hypothesis, which explored regional differences across the dependent measures of participant behavior, I also examined the possible relationship between three covariates and the same dependent measures. These covariates were participant performance, experience, and acuity. Just two MANCOVA tests examined many of the related covariates hypotheses (H1.2.1, H1.2.2, H2.1.1, H2.1.2, H2.3.1, and H2.3.2). These tests examined a multitude of different factors including all three covariates, three regional variables, and all proportions of evidence dependent variables. Due to the way the online A/B data system tracked users, I needed to use a separate MANCOVA to test for differences in the mouse movement variable. Find these tests in Appendix table 3.

The first covariate hypothesis considered the relationship between regional usage and spatial acuity. Specifically, there are two questions to consider:

> H2.1.1: The proportion of use for regions two and three will be significantly higher for individuals with higher spatial acuity.

and

> H2.1.2: The proportion of use for regions one, four, and five will be significantly lower for individuals with higher spatial acuity.

In examining the significance of the MANCOVA, time, distance, and velocity significantly varied in relation with all of the regional variables as well as two of the covariates: participant experience and participant acuity. However, participant performance was not significant within this model. The time spent between clicks was higher for regional groupings 1 and 2 as well as 1, 4, and 5. The LO acuity and LO experience group both had more time between mouse clicks, but the HI performance group had less. The distance between clicks was higher for regions 1 and 2 as well as for 2 and 3 and 4 and 5; groups containing region 1 had a smaller distance between clicks. The LO acuity, performance, and experience group all had more distance between mouse clicks. Velocity was lowest for groups containing region 3. The HI performance and experience groups moved more slowly, but the HI acuity group moved faster.

The next covariate to consider was experience. Two hypotheses related to this factor:

> H2.3.1: The proportion of use for regions four and five will be significantly higher for individuals who score HI on the experience with technology survey.

and

> H2.3.2: The proportion of use for regions one and two will be significantly lower for individuals who score HI on the experience with technology survey.

While there were slight differences in the mean time between mouse movements across all regional groupings, the MANOVA, which tested for differences between mouse movements, revealed that only movements from within regional grouping 2 and 3 significantly differed. In this case, compare 0.82 seconds between movements in region 2 and 3 to just 0.28 seconds for regions 1, 4, and 5. The HI acuity, performance, and experience groups all had more time between regional mouse movements.

Finally, hypothesis H2.2 considered whether performance related to experience:

> H2.2: Individuals in the HI performing group will have significantly higher overall scores on the experience with technology survey.

Because this experience survey had multiple questions, I used a MANOVA to examine interaction between participant performance group and the scores on this experience instrument. Overall, this model was not significant (Appendix table 4). A subsequent factorial ANOVA examined the interaction between performance and all eight questions on the experience instrument. No single factor related significantly to performance (Appendix table 5).

CHAPTER 5

DISCUSSION

The primary goal of the present study was to examine whether participants would use each of several sections of a dashboard differently. The secondary goal of the present study was to control for covariates such as performance, acuity, and experience that plausibly could affect participant usage patterns. Recall that there are four hypothesis. H0 relates to differences in usage patterns by region. The main effect section discusses this hypothesis since it is the primary goal of the present study. H1 considers whether participant performance interacts with any difference in behaviors across region. H2 considers differences in usage based upon spatial acuity. H3 considers whether experience interacts with participant usage patterns. The covariate effects section discusses H1, H2, and H3 since these sections relate to considering secondary effects.

Main effects

Whether or not participants could solve curricular decision making problems related to the way in which they used the dashboard to facilitate decision making. A number of factors are important to explaining the differences in the time between, distance between, and velocity of mouse clicks. Each regional grouping had a significantly different behavioral signature. However, the important regional groupings were different from expected. Of the exploratory behaviors that varied by region, distance traveled varied the most significantly.

Results related to H0

The primary hypothesis was that participants’ usage of the cognitive tool would vary by region.

> H0: Participants will use region 1 more than regions two and three; participants will use regions two and three more than regions four and five.

We can reject the null hypothesis for H0 that there would be no difference in usage patterns between regions. However the regional groupings are likely different than I anticipated. Based upon behavioral patterns, regional grouping [[1,2],[3,4],5] may better approximate actual cognitive functional groupings than the hypothesized grouping [1,[2,3],[4,5]]. The overall number of mouse clicks and mouse movements between regions is much higher for region 1 than for 2 and for 2 than for 3, 4, or 5. While these collections are different than I had originally anticipated, they do make sense when one considers the multiple stages of user interaction. Recall there is a three-stage model where stage 1 relates to region 1, stage 2 relates to regions 2 and 3, and stage 3 relates to regions 4 and 5. In order for a user to advance from one stage to another they would have interacted with the regions of the dashboard present within the previous stage. The following sections further explore the theoretical and performance based similarities between each of these new regional groupings. The following sections discuss results related to the new regional groupings described above. However, recall that the original hypotheses often focused upon the original set of regional groupings. For this reason, and in order to ease flow, I discuss some hypotheses multiple times.

Region 1 and 2 reduce cognitive load and provide scaffolding

I hypothesized that the overall activity level would be lower in regions one and two:

> H1.2.2: The proportion of use for regions one and two will be significantly lower within the higher performing group.

Regions 1 and 2 helped with exploration and with rapid interaction with the dashboard interface. Behaviors in these regions had longer durations between mouse clicks as well as relatively high velocities. This indicates that these regions are for rapid exploration. The distance between clicks is higher than average for region 2 but smaller than average for region 1 and likely relates to the difference in size of these two regions. The time between mouse movements across borders is much lower for region 1 suggesting perpetual revisiting of this region.

The revisiting behaviors within regions 1 and 2 are characteristic of cognitive tools intended to facilitate decision making through reducing cognitive load or through providing scaffolding. The research room and periodic table tools that Liu et al (2004) mention both served to offload cognitive load through concisely collocating relevant information. Similarly, regions 1 and 2 likely allowed participants to reduce the complexity of the decision space. The launcher room tool helped students design missions using various budget constraints. In a similar fashion, regions 1 and 2 illustrated the relationship and relative performance of concepts within the ISUCVM ontology. The information in regions 1 and 2 served as a building block for future participant interactions in the subsequent dashboard regions.

Regions 3 and 4 provide scaffolding and support attribution formation

I hypothesized that the overall activity level would be higher in region four, however I did not have a specific hypothesis linked to activity levels for region 3:

> H1.2.1: The proportion of use for regions four and five will be significantly higher within the higher performing group.

Regions 3 and 4 are more highly interactive and likely assist in hypothesis formation. Participants used region 3 more often than region 4. Participants used these regions less often and in more detail than they used regions 1 and 2. There was less time between mouse clicks than average. The distance traveled between clicks was similar to average. However, the velocity was slower than average. Since time was faster, velocity was slower, and distance was about the same then this combination of results suggests that the interaction within these regions was more interactive.

The higher continuity of movements in regions 3 and 4 are evidence that participants' behaviors are migrating towards a solution to a particular problem. Similar to regions 1 and 2, regions 3 and 4 also provide scaffolding. The scaffolding comes from the contextual evidence that aggregates related, in-depth performance data. However, here the interaction between the participants and these regions directly facilitates attribution formation in ways that regions 1 and 2 do not. For example, the control room in the study by Liu et al (2004) facilitated student attribution by providing the results of past exploration in a single location. Similarly, regions 3 and 4 show the in-depth performance results related to the concepts in regions 1 and 2. Region 4 contains information related to the context for each concepts. Region 3 provided event-related meta-data to more directly help participants to attribute concepts. At the time of the study, this section only contained a few meta-data concepts. In the future, this section will contain many more meta-data concepts.

Region 5 facilitates cognitive process

Recall that I hypothesized that the overall activity level would be higher in region five:

> H1.2.1: The proportion of use for regions four and five will be significantly higher within the higher performing group.

As it turned out, participants rarely used region 5. Peculiarly, this region had the shortest time between clicks of all regions. It also had the largest distance traveled and the highest velocity between mouse clicks. Participants rapidly explored this section searching for details about concepts in which they were already interested. Based upon interaction patterns, participants used this section for more content directed, in-depth exploration.

This section provides extended data for historic events related to the concepts present within the previous four sections. Because there are multiple stages of interaction within the tool, this means that participants would have already had to explore these concepts in the other regions in order to see the information within this section. Because the access of information in this section requires users to engage in a multi-stage flow of interactions, this region facilitates decision making through facilitating the cognitive process. The function of this section is similar to the way that the notebook and the bookmark tool from Liu et al (2004) study allow students to assemble together information from previous interactions.

Covariate effects

As mentioned I used a single MANCOVA to examine whether behaviors and covariate participant factors significantly varied by region. Based upon this overall model, patterns of use depended upon the spatial acuity and the experience of the participant. Performance was not significantly explanatory to this overall model. However, subsequent tests found slightly different results. A MANOVA that specifically examined experience found that experience had no relationship with either performance or any of the other exploratory variables. While performance was not significant within the overall model, it was significant when examined on its own. The following subsections discuss each covariate in more depth.

The differences in patterns of use between regions are not due to differences in participant experience with technology or to experience with making curricular decisions. However, these differences may be due to either or both of the two covariates: performance and spatial acuity. The levels of both of these covariates varied alongside differences within the dependent measures.

While never specifically tested as a part of the hypotheses, the relationship between performance and acuity is unclear. Because I measured acuity independent from performance using the embedded figures test prior to beginning the experiment, it is possible to claim that acuity influences performance. However, since I had no similar pre-experimental measure of performance, I am not sure whether performance influenced acuity.

Performance

I thought performance would interact both with the way participants used the dashboard and with participant experience with technology. Two hypotheses addressed these questions:

> H1.1: The higher performing group will have a lower overall number of exploratory behaviors and of number of concepts considered.

and

> H2.2: Individuals in the HI performing group will have significantly higher overall scores on the experience with technology survey.

As mentioned, when considered amongst the other covariates experience and spatial acuity, performance does not significantly contribute to the behavioral model. However, on its own performance was significant within the behavioral model. When examined individually, only the velocity and distance significantly varied. The LO performance group traveled a significantly further distance than the HI performance group and at a significantly slower speed. The HI performance group exhibited considerably slower exploratory behaviors. Nonetheless, the two types of evidence, the time between regional movements (0.77 seconds HI v. 0.55 seconds LO) and number of concepts considered (42.87 HI v. 42.52 LO), did not vary due to performance. Those with HI performance are not necessarily those with HI levels of experience with technology or those with curricular decision making experience. Instead, those who used the dashboard more thoroughly made better decisions.

Experience

Two hypotheses considered whether a participants' experience with technology would interact with their usage patterns:

> H2.3.1: The proportion of use for regions four and five will be significantly higher for individuals who score HI on the experience with technology survey.

and

> H2.3.2: The proportion of use for regions one and two will be significantly lower for individuals who score HI on the experience with technology survey.

While experience was significant within the overall model, it was not significant when examined on its own. Participant experience with cognitive tools and with technology did not directly relate to differences in usage patterns. However, it is possible that experience acts as a covariant alongside participant behaviors.

Spatial acuity

I thought spatial acuity would affect the use of various dashboard regions. Two hypotheses considered this question:

> H2.1.1: The proportion of use for regions two and three will be significantly higher for individuals with higher spatial acuity.

and

> H2.1.2: The proportion of use for regions one, four, and five will be significantly lower for individuals with higher spatial acuity.

Success, as measured by performance, relates to using the dashboard regions for the correct purpose. Individuals with higher spatial acuity have a better knack for figuring this out on their own. The behavioral patterns of the HI v. LO spatial acuity group were significantly different. The HI group spent 186% as much time in between mouse clicks as the LO group but only moved 80% of the distance. The HI group also spent 23% more time in between mouse movements than the LO group. However, the HI spatial acuity group was not the same as the HI performance group. Spatial acuity may intervene with performance through effecting the way that users interact with the dashboard.

CHAPTER 6

CONTRIBUTION

The information visualization platform created for use within this study facilitates decision making independent of experience. Participants with higher overall usage behaviors tended to have better curricular decision making outcomes. This effect was independent of experience but it may vary due to spatial acuity.

This study examines previous research on cognitive tools and defines various types of cognitive tools based upon how they facilitate decision making. This study then uses participants' behavioral patterns to classify each of several regions within an interface as belonging to one or several types of cognitive tools. This quantitative approach allows for a finessed examination of the usability of particular areas of an interface based upon their structural and functional capabilities. This has implications for making informed recommendations for design changes where the goal is to facilitate certain cognitive capabilities.

This study also supports previous research showing the capability of using just mouse tracking for informing design decisions. Here mouse tracking was effective for tracking low involvement website browsing behavior. Furthermore, approaches such as mouse tracking allow for more rapid development of online testing frameworks. While this study examined participants for a set duration within a focus group setting, this study used several forms of online user and mouse tracking to record participants' exploratory behaviors. This study extends the use of these online tracking techniques to the realm of high-involvement decision making using an information visualization platform. The potential for the use of this approach promises to create less intrusive, more naturalistic experiments.

CHAPTER 7

LIMITATIONS

This study has theoretical and applied limitations. Although the results for differences in regional usage were significant, the sample size for this study was, for practical reasons, constrained. For this reason, much of the statistical power for this study came from measuring large quantities of behavior for a small set of participants. This means that, although there is empirical, statistical power, the interpretation of these results may not be generalizable to other users or to a broad spectrum of cognitive tool contexts.

In addition to applied limitations, this study also relies upon the assumption that participants' mouse behaviors can serve as a proxy for their thoughts or for their intentions. While there is evidence of a link between mouse movements and eye movements, this link can only account for about 75% of the mouse movements (Chen, Anderson, & Sohn, 2001). Furthermore, studies that use eye tracking assume that what people fixate upon represents that which they are thinking. While this study found significant differences within participant behaviors, this assumption could complicate the conclusion that these differences represent differences in cognitive function.

Finally, certain issues related to scales could complicate the use of the MANCOVA, MANOVA, and ANOVA tests that I employed. For example, I did not normalize or test for the goodness of fit to a normal distribution for the dependent variables. I also did not perform these tests for participants' scores on the experience with technology surveys. While the latter represents a set of scales that are arguably criterion-based, the former represents a set of behaviors that could potentially vary nonlinearly based upon individual differences between participants. For this reason, further research needs to validate whether the natural variation in participant mouse movements (eg. distance traveled, speed of movement, and time between clicks) need statistical adjustment prior to analysis.

CHAPTER 8

NEXT STEPS

Using the results of this research as a baseline, the next step will be to examine user behaviors over time. Given the results of this study, these behaviors could be studied using abbreviated measures. It is possible to put ambient user testing in place that allows continued examination of mouse and concept tracking systems. This would allow for continued ad hoc usability analysis as well as a longitudinal study of the effectiveness of implementing this cognitive tool within the related ISUCVM context.

Future research should consider controlling for two additional factors. First, there is some evidence even within this study that motivation and or interest effect the performance of participants. While it is doubtful that any participants were carefree, there are likely differences in the level of personal involvement each feels with the decision making problem at hand. Stratifying participants based upon an attitudinal or interest survey would make it possible for myself and for others to test this in the future. Second, examine whether the effect of spatial acuity upon performance and upon exploration style interacts with learning. Accomplish this through doing away with the counterbalancing technique used in this study. Whereas counterbalancing helps normalize the learning effect across groups, it also prevents analysis from considering the interaction between this effect and other variables or covariates.

REFERENCES

Atkinson, R. C. & Shiffrin, R. M. (1971). The control of short-term memory. *Scientific American, 225(2),* 82-90.

Ayed, M.B., Ltifi, H., Kolski, C. & Alimi, A.M. (2010). A user-centered approach for the design and implementation of KDD-based DSS: A case study in the healthcare domain. *Decision Support Systems 50(2010),* pp. 64-78.

Baddeley, A. (2002). Is working memory working? *European Psychologist,* 85-97.

Bargh (1992). The ecology of automaticity: Toward establishing the conditions needed to produce automatic processing effects. *The American Journal of Psychology, 105,* 181-199.

Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning.* New York, NY: McGraw-Hill.

Bowers, A. J. (2009). Reconsidering grades as data for decision making: More than just academic knowledge. *Journal of Educational Administration, 47(5),* 609-629.

Brown, S. C., & Craik, F. I. M. (2000). Encoding and retrieval of information. In E. Tulving & F. I. M. Craik (Eds.) *The Oxford handbook of memory* (pp. 93-107). New York: Oxford University Press.

Chen, M.-C., Anderson, J. R., & Sohn, M.-H. (March 31, 2001). What can a mouse cursor tell us more? Correlation of eye/mouse movements on web browsing. *Conference on Human Factors in Computing Systems.*

Chepenik, L. G., Cornew, L. A., & Farah, M. J. (2007). The influence of sad mood on cognition. *Emotion, 7,* 802-811.

Collins, A. M. & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82(6),* 407-428.

Craik, F.I.M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104,* 268-294.

Crain, W. (1985). *Theories of development: Concepts and applications* (2nd ed.). Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Danielson, J.A., Mills, E.M, Vermeer, P.J., Preast, V.A., Young, K.M., Christopher, M.M., George, J.W., Wood, D.R. and Bender, H.S. (2007). Characteristics of a Cognitive Tool that Helps Students Learn Diagnostic Problem Solving. *Educational Technology Research and Development, 55,* 499-520.

Danielson, J.A., Mills, E.M., Vermeer, P.J., and Bender, H.S. (2008). The Diagnostic Pathfinder: Ten years of using technology to teach diagnostic problem solving. In F. Columbus (Ed.), *Educational technology research advances.* Nova Science Publishers.

Driver, J. (2001). A selective review of selective attention research from the past century. *British Journal of Psychology, 92,* 53-78.

Drummond, H. (1995), De-escalation in decision making: A case of disastrous partnership. *Journal of Management Studies, 32,* 265–281.

Egan, D. E. (1988). Individual differences in human-computer interaction. In Helander, M. (Ed.), Handbook of human-computer interaction. Elsevier Science Publishing Company, Inc., New York.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*, 19-23.

Fernandez-Duque, D. & Johnson, M. L. (2002). Cause and effect theories of attention: The role of conceptual metaphors. *Review of General Psychology, 6*, 153-165.

Fitzpatrick, J.L., Sanders, J.R., & Worthen, B.R. (2004). *Program evaluation: Alternative approaches and practical guidelines (3rd ed.).* New York, NY: Longman.

Foer, J. (2011). Moonwalking with Einstein: The art and science of remembering everything. The Penguin Press: New York.

Frederiksen, C. Kehoe, J. E. & Wood, R. (2011). Effects of instructional aids on the acquisition of dynamic decision making skills. *Learning and Instruction 21(2011)*, pp. 601-613.

Gong, Y., & Zhang, J. (2011). Toward a human-centered hyperlipidemia management system: The interaction between internal and external information on relational data search. *J. Med. Syst.35(2011)*, pp. 169-177.

Greenwald. (1992). New look 3: Unconscious cognition. *American Psychologist, 47,* 766-779.

Holloway, R. E. (2002). Diffusion and adoption of educational technology: A critique of research design. *Educational Communications and Technology*. Presented at the Association for Educational Communications and Technology.

Holyoak, K. J. (2008). Relations in semantic memory: Still puzzling after all these years. In M. A. Gluck, J. R. Anderson, & S. K. Kosslyn (Eds.), *Memory and mind: A festschrift for Gordon A. Bower* (pp. 141-158). New York: Erlbaum.

Janis, I.L. (1971). Groupthink. *Psychology Today 5(6)*, 43-46.

Jarupathirun, S. & Zahedi, F. M. (2007). Exploring the influence of perceptual factors in the success of web-based spatial DSS. *Decision Support Systems 43(2007)*, pp. 933-951.

Jonassen, D. (2003). Using cogntivie tools to represent problems. *Journal of Research on Technology in Education, 35*, 36-82.

Jones, E. E. & Davis, K. E. (1965). *From acts to dispositions: The attribution process in person perception.* In G. B. Moskowitz, *Social Cognition* (pp. 268-270). New York: Guilford Press.

Kim, B. & Reeves, T.C. (2007). Reframing research on learning with technology: In search of meaning of cognitive tools. *Instructional Science 37(2007)*¸207-256.

Klatzky, R. L. (1980). *Human memory: Structures and processes* (2nd ed., pp. 236-254). San Francisco: Freeman.

Knapp, M.S., Copland, M.A. & Swinnerton, J.A. (2007). Understanding the promise and dynamics of data informed leadership. In P. Moss (Ed.), *Evidence and decision making* (Vol. 106, No. 1: National Society for the Study of Education yearbook, pp. 15-45). Malden, MA: Blackwell Publishing.

Kosslyn, S. M., Thompson, W. L., & Ganis, Giorgio (2002). Mental imagery doesn't work like that. (Commentary on target article by Z. W. Pylyshyn, Mental imagery: In search of a theory). *Behavioral and Brain Sciences, 25,* 198-200.

Lajoie, S. P. (1993). Computer environments as cognitive tools for enhancing learning. In S. P. Lajoie and S. J. Derry (Eds.), Computers as cognitive tools (pp.261–288). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Lankton, N. K., Speier, C., & Wilson, V.E. (2012). Internet-based knowledge acquisition: Task complexity and performance. *Decision Support Systems 53(2012),* pp. 55-65.

Lazar, J., Feng, J. H. & Hochheiser, J. (2010). *Research methods in human-computer interaction.* Chichester, West Sussex, U.K.

Liu, M., & Bera, S. (2005). An analysis of cognitive tool use patterns in a hypermedia learning environment. Educational Technology Research and Development, 53(1), pp 5-21.

Liu, M., Bera, S., Corliss, S. B., Svinicki, M. D., & Beth, A. D. (2004). Understanding the connection between cognitive tool use and cognitive processes as used by sixth graders in a problem-based hypermedia learning environment. Journal of Educational Computing Research, 31(3), pp 309-334.

Liu, M., et al (2009). Students' problem solving as mediated by their cognitive tool use: A study of tool use patterns. J. Educational Computing Research, 40(1), pp. 111-139

McClelland, J. L. (2000). Connectionist models of memory. In E. Tulving & F. I. M. Craik (Eds.). *The Oxford handbook of memory* (pp. 583-596). New York: Oxford University Press.

Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist, 44,* 1469-1481.

Mertler, C. A. & Vanatta, R. A. (2010). Advanced and multivariate statistical methods: Practical application and interpretation (4th Ed.). Pyrczak Publishing.
Mirolli, M. & Parisi, D. (2009). Language as a cognitive tool. *Minds & Machine 2009(19),* 517 – 28.

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin,132*, 297-326.

Moss, P. (2007). Introduction: Evidence and decision making. In P. Moss (Ed.), *Evidence and decision making* (Vol. 106, No. 1: National Society for the Study of Education yearbook, pp. 15-45). Malden, MA: Blackwell Publishing.

Oltman, P. K, Raskin, E. & Witkin, H. A. (1971). Group embedded figures test. Consulting Psycholgits Press, Inc.

Parker, G. (2012). Acta is a four-letter word. *Acta Psychiatrica Scandinavia, 126(6),* 476-478.

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Pedhazur, E. J. & Schmelkin, L. P. (1991). Measurement, Design and Analysis: An integrated approach. Lawrence Erlbaum Associates, Publishers: Hillsdale, NJ.

Preast, V.A., Danielson, J.A., Bender, H.S. and Bousson, M. (2007). Effectiveness of a computer-based tutorial for teaching how to make a blood smear. *Veterinary Clinical Pathology, 36(3),* 245-252.

Pylyshyn, Z. W. (2003b). Return of the mental image: Are there really pictures in the brain? *Trends in Cognitive Sciences, 7,* 113-118.

R.D. Johnson, G.M. Maraka (2000). Research report: the role of behavioral modeling in computer skills acquisition toward refinement of the model. *Information Systems Research 11 (4),* pp. 402–417.

Randall, D., Harper, R. Rouncefield, M. (2007). *Fieldwork for design: Theory and practice.* Springer-Verlag: London.

Roediger, H. L., III, McDermott, K. B. (2000). Tricks of memory. *Current Directions in Psychological Science, 9,* 123-127.

Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). New York: The Free Press.
Salomon, G. (1988). AI in reverse: Computer tools that turn cognitive. J. Educational Computing Research, 4(2), 123–139.

Salomon, G., Goberson, T. & Guterman, E. (1981). The computer as a zone of proximal development: Internalizing reading-realted metacognitions from a reading partner. *Journal of Educational Psychology, 81(4)*, 620-627.

Shiffrin, R. M. & Schneider, W. (1977). Controlled and automatic human information processing: II Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127-190.

Shiffrin, R. M. (1993). Short-term memory: A brief commentary. *Memory & Cognition, 21*, 193-197.

Skinner, B. F. (1987). Whatever happened to psychology as the science of behavior. *American Psychologist, 42,* 780-786.

Smith, E. E., Kossylyn, S. M. (2007). *Cognitive psychology: Mind and Brain.* Upper Saddle River, NJ: Pearson Prentice Hall.

Sohn, M.-H., Douglass, S.A., Chen, M.-C., & Anderson, J.R. (2005). Characteristics of fluent skills in a complex, dynamic problem-solving task. *Human factors: The journal of the human factors and ergonomics society.*

Speier,C., Valacich, J.S., & Vessey,I. (1999). The influence of task interruption on individual decision making: an information overload perspective. *Decision Sciences 30 (2),* pp. 337–360.

Stemler, Steven E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation, 9(4)*. Retrieved May 7, 2012 from http://PAREonline.net/getvn.asp?v=9&n=4

Stroemeyt, N., Franks, N. R. & Giurfa, M. (2011). Knowledgeable individuals lead collective decisions in ants. *The Journal of Experimental Biology 214*, pp. 3046-3054.

Sukhoy, V. and Stoytchev, A. (2010, December 6-8). Learning to detect the functional components of doorbell buttons using active exploration and multimodal correlation. *Proceedings of the 10th IEEE International Conference on Humanoid Robots.* Paper presented at Humanoids, Nashville, TN.

Swan, G., & Mazur, J. (2011). Examining data driven decision making via formative assessment: A confluence of technology, data interpretation heuristics and curricular policy. *Contemporary Issues in Technology and Teacher Education, 11(2),* 205-222.

Townsend, A. M. & Bennet, J. T. Privacy, technology and conflict: Emerging issues and action in workplace privacy. In: Taras, D.G., Bennet, J.T. & Townsend, A.M. (2007). *Information Technology and the world of work* (pp. 191-202). London: Transaction Publishers.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology, 53,* 1-25.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, New Series, 185(4157),* 1124-1131.

Tversky, A. & Kahneman, D. (1981). The framing of decision and the psychology of choice. *Science, New Series, 211(4481),* 453-458.

Vygotsky, L.S. (2010). Consciousness as a problem in the psychology of behavior. *Russian Social Science Review [1061-1428] Vygotsky, 20(4)*, 47-79.

Wayman, J. C. (2005). Involving teachers in data-driven decision making: Using computer data systems to support teacher inquiry and reflection. *Journal of Education for Students Placed at Risk*, *10(3)*, 295-308.

Wexler, B. E. (2006). *Brain and culture.* London: MIT Press.

Yalamova, R. (2009). Correlations in financial time series during extreme events-spectral clustering and partition decoupling method. *Proceedings of the World Congress on Engineering*, 2. London, U.K.

Xu, L. (1997). Bayesian Ying-Yang machine, clustering and number of clusters. *Pattern Recognition Letters 18(1997)*. Pp. 1167-78.

Xu, L. (2010). Bayesian Ying-Yang system, best harmony learning, and five action circling. *Front. Electr. Electron. Eng. China 5(3)*. pp 281-328.

APPENDIX A

INSTRUMENTS

Tasks

The tasks asked about these three areas: Basic Science, Clinical Competency, and College. These three areas of the dashboard have the most complete data. The College area does not relate to the domain of veterinary medical knowledge, but relates to overall satisfaction with collegiate facilities. The general template follows this format: "In the area of [area], what are the problem areas and the areas of most success and what is your top recommendation for change? Please explain your answers." When participants responded, they had answer space to answer each of six implied questions for each task:

- Problem area?
- Please explain why you chose those problem areas.
- Areas of most success?
- Please explain why you chose those areas of success.
- Please indicate your top recommendation for change.
- Please explain why you selected this change

Rubric

A panel of experts used this rubric to grade the quality of each participant's response to each of three questions. The figure below reproduces this panel as shown to these experts.

	Really good	Good	Neither good nor poor	Poor	Really poor
Plausibility of the solution	The solution answers the questions posed in the problem and is compelling.	The solution could explain the questions posed in the problem.	It is unclear whether the solution answers the questions posed in the problem.	The solution probably does not answer the questions posed in the problem.	The solution definitely does not answer the questions posed in the problem.
Comprehensiveness of the solution	The solution addresses all relevant issues that could be associated with the problem.	The solution addresses some of the relevant issues associated with the problem, but leaves some relevant issues unaddressed.	The solution addresses some relevant issues, but leaves important relevant issues unaddressed, and/or addressses unrelated issues.	The solution leaves many relevant issues unaddressed and/or it also addresses a number of unrelated issues.	The solution does not appear to address any of the relevant issues.
Optimality of the solution	The solution is as efficient as possible.	The solution is quite efficient, though more efficient alternatives exist.	The solutions is of average efficiency. Alternatives that are both more and less efficient exist.	The solution is likely to work, but in not very efficient in comparison with alternatives.	The solution is so inneficient that it would probably not work.
Solution supported by evidence	The evidence supports the solutions, with no flaws in logic or reasoning.	The evidence generally supports the solution, with some minor flaws in logic or reasoning.	The evidence generally upports the solution, though there is at least one important flaw in logic or reasoning.	The evidence provides some support for the solution, though there are several significant flaws in logic or reasoning.	The solution is illogical and does not follow from the evidence cited.
Comprehensiveness of the evidence considered	The individual considered sufficient evidence to produce a good solution.	The indvidual considered relevant evidence for supporting the solution, though some minor relevant evidence was ommitted.	The individual considered some relevant evidence, while ommitting some important evidence.	The individual considered very little relevant evidence, impacting the quality of the solution.	The individual considered so little relvant evidence that it was impossible to produce a reasonable solution.

Experience Survey

Consider this sample item:

	Agree?	Not at all confident				Totally confident
I believe I have the ability to use a website.	**YES** NO	1	2	3	4	5

Please read each question and answer based only upon your personal feelings.

	Agree?	Not at all confident				Totally confident
I believe I have the ability to navigate unfamiliar websites.	YES NO	1	2	3	4	5
I believe I have the ability to manipulate the way data appears within an online dashboard (interactive website).	YES NO	1	2	3	4	5
I believe I have the ability to summarize numerical information that appears in online websites.	YES NO	1	2	3	4	5
I believe I have the ability to navigate lists that can be collapsed and expanded.	YES NO	1	2	3	4	5
I believe I have the ability to navigate unfamiliar graphs.	YES NO	1	2	3	4	5
I believe I have the ability to summarize unfamiliar graphs.	YES NO	1	2	3	4	5
I believe I have the ability to determine relative value for each graph in a series of graphs without examining each in detail.	YES NO	1	2	3	4	5
I believe I have the ability to use websites to assist in making decisions.	YES NO	1	2	3	4	5
I believe I have the ability to form decisions about curricular change.	YES NO	1	2	3	4	5

Based upon Marakas, Johnson, and Yi (1999).

APPENDIX B

TABLES RELATED TO RESULTS

Table 1: Tests Related to H0

MANCOVA for Click Data

	Λ	F	df1	df2
Clicks by Region	1	8.95***	4.00	930

ANOVA for Mouse Events Data

Source	df	F	η	p
Events by Region	4	3.22*	0.03	0.01
error	11060	(0)		

ANOVA for Time between Mouse Clicks

Source	df	F	η	p
Time	4	1.49	0.08	0.20
error	930	(8.83)		

ANOVA for Distance between Mouse Clicks

Source	df	F	η	p
Distance	4	21.17***	0.30	0.00
error	930	(49.7)		

ANOVA for Velocity of Mouse Movements between Clicks

Source	df	F	η	p
Velocity	4	1.32	0.08	0.26
error	930	(47.08)		

. *$p < 0.10$, * $p<0.05$, ** $p < 0.01$, *** $p<0.001$*

Average Click Behaviors by Region

Region	Sample Size	Time (sec.)	Distance (pixels)	Velocity (pixel/sec.)
1	447	28.76	84.14	50.62
2	273	24.03	228.6	48.31
3	127	9.07	86.59	11.91
4	81	14.14	155.2	17.14
5	7	2.74	210.00	97.39
All	935	23.24	133.74	42.14

Average Events by Region		
Region	Mean	Sample Size
1	0.26	2410
2	0.80	7617
3	1.11	566
4	0.32	241
5	0.57	231
All	0.68	11065

Table 2: Tests Related to H1

MANCOVA for Click Data

	Λ	F	df1	df2
Clicks by Region	1	4.46**	1.00	933

ANOVA for Number of Mouse Events by Performance

Source	Df	F	H	p
Events by Performance	1	2.4	0.01	0.12
error	11063	(0)		

ANOVA for Number of Concepts Considered by Performance

Source	Df	F	H	P
Concepts by Performance	1	0	0.01	0.96
error	88	(14.14)		

ANOVA for Time between Mouse Clicks

Source	Df	F	H	P
Time	1	0.81	0.03	0.37
error	933	(8.82)		

ANOVA for Distance between Mouse Clicks

Source	Df	F	H	P
Distance	1	4.84*	0.07	0.03
error	933	(53.6)		

ANOVA for Velocity of Mouse Movements between Clicks

Source	Df	F	H	P
Velocity	1	6.85**	0.09	0.01
error	933	(46.7)		

. *p < 0.10, * p<0.05, ** p < 0.01, *** p<0.001*

Average by Performance

Group	Sample Size	Time (sec.)	Distance (pixels)	Velocity (pixel/sec.)
Hi	631	25.10	122.6	29.74
Lo	304	19.40	156.9	67.87
All	935	23.24	133.74	42.14

Average by Performance

	Mean		Sample Size	
Level	Concepts	Events	Concepts	Events
Hi	42.87	0.77	46.00	6788
Lo	42.52	0.55	44.00	4277
All	42.70	0.68	90.00	11065

Table 3: Tests Related to H1.2.1, H1.2.2, H2.1.1, H2.1.2, H2.3.1, and H2.3.2

MANCOVA for Click Data by Region by Performance by Experience

	Λ	F	df1	df2
Clicks by Regions (1,2)	0.99	3.37*	3	926
Clicks by Regions (2,3)	0.95	17.7***	3	926
Clicks by Regions (4,5)	0.95	15.55***	3	926
Clicks by Acuity	0.98	6.06***	3	926
Clicks by Performance	1.00	0.98	3	926
Clicks by Experience	0.98	6.2***	3	926

Factorial ANOVA for Events Data by Region by Performance by Experience

	Df	F	η	P
Events by Regions (1,2)	1	0.33	0.00	0.57
Events by Regions (2,3)	1	12.35***	0.00	0.00
Events by Regions (4,5)	1	0.07	0.00	0.79
Events by Acuity	1	1.04	0.00	0.31
Events by Performance	1	1.07	0.00	0.30
Events by Experience	1	0.74	0.00	0.39
error	11058	(0.0047)		

. *p < 0.10, * p<0.05, ** p < 0.01, *** p<0.001*

Average Time between Events by Region Group

Regions	Mean	Sample Size
3,4,5	0.80	1038
1,2	0.67	10030
All	0.68	11068

Average Time between Events by Region Group

Regions	Mean	Sample Size
1,4,5	0.28	2882
2,3	0.82	8183
All	0.68	11065

Average Time between Events by Region Group

Regions	Mean	Sample Size
1,2,3	0.68	10590
4,5	0.62	472
All	0.68	11062

Average Events by Group

	Mean			Sample Size		
Level	Acuity	Performance	Experience	Acuity	Performance	Experience
Hi	0.74	0.72	0.73	6165	6788	*6324*
Lo	0.60	0.62	0.62	4900	4277	*4741*

Average Time by Region Group

Regions	Mean	Sample Size
3,4,5	10.77	215
1,2	26.97	720
All	23.24	935

Average Time by Region Group

Regions	Mean	Sample Size
1,4,5	25.15	535
2,3	20.7	400
All	23.24	935

Average Time by Region Group

Regions	Mean	Sample Size
1,2,3	23.27	847
4,5	22.99	88
All	23.24	935

Average Time by Group

	Mean			Sample Size		
Level	Acuity	Performance	Experience	Acuity	Performance	Experience
Hi	28.97	21.60	30.29	535	631	*576*
Lo	15.59	26.66	11.94	400	304	*359*

Average Distance by Region Group

Regions	Mean	Sample Size
3,4,5	116.5	215
1,2	138.9	720
All	133.74	935

Average Distance by Region Group

Regions	Mean	Sample Size
1,4,5	95.08	535
2,3	185.4	400
All	133.74	935

Average Distance by Region Group

Regions	Mean	Sample Size
1,2,3	123.5	847
4,5	232	88
All	133.74	935

Average Distance by Group

	Mean			Sample Size		
Level	Acuity	Performance	Experience	Acuity	Performance	Experience
Hi	123.50	131.10	127.20	535	631	*576*
Lo	147.50	139.30	144.30	400	304	*359*

Average Velocity by Region Group

Regions	Mean	Sample Size
3,4,5	16.66	215
1,2	49.74	720
All	42.14	935

Average Velocity by Region Group

Regions	Mean	Sample Size
1,4,5	44	535
2,3	39.65	400
All	42.14	935

Average Velocity by Region Group

Regions	Mean	Sample Size
1,2,3	41.7	847
4,5	46.34	88
All	42.14	935

Average Velocity by Group

	Mean			Sample Size		
Level	Acuity	Performance	Experience	Acuity	Performance	Experience
Hi	28.97	38.33	41.23	535	631	*576*
Lo	15.59	50.03	43.58	400	304	*359*

Table 4: Tests Related to H2.2

MANOVA for Click Data by Region by Performance by Experience

	Λ	F	df1	df2
Clicks by Region	1	0.96	1.00	10

ANOVA for Click Data by Region by Performance by Experience for Ability To Navigate Unfamiliar Interfaces

Source	df	F	H	p
Navigate Unfamiliar Interfaces	1	0.91	0.30	0.36
Error	10	(0.04)		

ANOVA for Click Data by Region by Performance by Experience for Ability To Manipulate Data Appearance

Source	df	F	η	p
Manipulate Data Appearance	1	0.92	0.30	0.36
Error	10	(0.08)		

ANOVA for Click Data by Region by Performance by Experience for Ability To Summarize Numerical Information

Source	df	F	η	p
Summarize Numerical Information	1	1.43	0.38	0.26
Error	10	(0.09)		

ANOVA for Click Data by Region by Performance by Experience for Ability To Navigate Trees

Source	df	F	η	p
Navigate Trees	1	4.5.	0.67	0.06
Error	10	(0.07)		

ANOVA for Click Data by Region by Performance by Experience for Ability To Navigate Graphs

Source	df	F	η	p
Navigate Graphs	1	0.23	0.15	0.64
Error	10	(0.14)		

ANOVA for Click Data by Region by Performance by Experience for Ability To Relative Graph Interpretation

Source	df	F	η	p
Relative Graph Interpretation	1	0.09	0.10	0.77
Error	10	(0.09)		

ANOVA for Click Data by Region by Performance by Experience for Ability To Online Research

Source	df	F	η	p
Online Research	1	0.63	0.25	0.45
Error	10	(0.05)		

ANOVA for Click Data by Region by Performance by Experience for Ability To Curricular Change

Source	df	F	η	p
Curricular Change	1	0	0.00	1.00
Error	10	(0.08)		

Table 5: Participant Groups

Average Acuity Score by Acuity Level

	Section I	Section II
Hi	6.83	8.00
Lo	3.00	3.50
All	4.92	5.75

Average Performance Score by Performance Level

	Comprehensive Evidence	Comprehensive Solution	Optimal Solution	Plausible	Supporting Evidence
Hi	3.39	3.18	3.21	3.70	3.61
Lo	1.84	1.86	1.85	2.21	2.00
All	2.61	2.52	2.53	2.96	2.80

Average Experience Survey Score by Experience Level

Regions	Navigate Unfamiliar Interfaces	Manipulate Data Appearance	Summarize Numerical Info.	Navigate Trees	Navigate Graphs	Relative Graph Interpretation	Online Research	Curricular Change
Hi	4.17	3.67	3.67	4.67	3.50	3.67	4.00	3.83
Lo	3.83	3.17	3.00	3.67	3.17	3.50	3.67	3.83
All	4.00	3.42	3.33	4.17	3.33	3.58	3.83	3.83

BIOGRAPHICAL SKETCH

Ryan Allen Kirk was born in Minneapolis, Minnesota. He attended Coon Rapids High School and received a B.S. in Psychology from Drake University in 2009. He served as a research assistant at Iowa State University as a part of the Office of Curricular and Student Assessment at the College of Veterinary Medicine.